DOCUMENT RESUME

ED 151 522 CE 014 935

AUTHOR Schwind, Hermann F.
TITLE A New Behavior Oriented Evaluation Scale to Assess Teaching and Training Effectiveness.
PUB DATE Oct 77
NOTE 30p.; Paper presented at the Evaluation Research Society Meeting (Washington, D.C., October 1977). For a related document see ED 141 507

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Administrative Personnel; Administrator Evaluation; *Behavior Rating Scales; Employees; Employment Qualifications; Information Needs; Information Systems; Job Analysis; Job Skills; Management Information Systems; Management Systems; Measurement Instruments; Measurement Techniques; *Personnel Evaluation; Rating Scales; Task Analysis; *Task Performance; Test Reviews; *Test Validity
IDENTIFIERS Behaviorally Anchored Rating Scale; *Behavior Descriptor Index

ABSTRACT

The first of several on-going tests was conducted of the Behavior Descriptor Index (BDI) which is a new behaviorally oriented job performance appraisal and training evaluation instrument developed for five large Canadian chartered banks. The new instrument was shown to be valid across organizations. Twenty-five assistant bank managers in these five banks were evaluated by their superiors using the BDI as well as the Graphic Rating Scale and the Behaviorally Anchored Rating Scale (BARS). Three conclusions were drawn from the results of this study. First, behaviorally oriented scales need not be organization specific. Second, it is possible to improve the psychometric characteristics of behavior oriented scales by switching from a Likert-type scale with specific anchor points to an anchorless format which allows the use of the total job behavior domain. And last, it is possible to use BDI information as a basis for a personnel management system because of its favorable rating characteristics, the increased utilization of job information, and the very specific feedback to raters, evaluators, and ratees. (EM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED151522

A NEW BEHAVIOR ORIENTED EVALUATION SCALE TO ASSESS TEACHING AND TRAINING EFFECTIVENESS

Paper presented at the
Evaluation Research Society Meeting in
Washington, D.C., October 1977

by

Hermann F. Schwind
Faculty of Commerce
Saint Mary's University
Halifax, Canada B3H 3C3

CE 014 935

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Herman F. Schwind

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ACKNOWLEDGMENT

The author would like to thank the Institute of Canadian Bankers, especially its Deputy Director for Research and Development, Mr. Peter H. Wigram, and the executives of the Royal Bank of Canada, the Canadian Imperial Bank of Commerce, the Bank of Montreal, the Toronto Dominion Bank, and the Bank of Nova Scotia for their generous support of the project.

EVALUATION RESEARCH SOCIETY MEETING
Washington, D.C.
October 1977

Hermann F. Schwind
Faculty of Commerce
Saint Mary's University
Halifax, CANADA B3H 3C3

ABSTRACT

A new behaviorally oriented performance appraisal and training evaluation instrument was developed for five large Canadian chartered banks. The new instrument has the advantage that it is valid across organizations. Twenty-five assistant bank managers were evaluated by their superiors, using three different appraisal instruments: Graphic Rating Scale, Behaviorally Anchored Rating Scale, and Behavior Description Index. Campbell and Fiske's Multitrait-multimethod matrix was used to test the construct validity of the three tests employed. The Behavior Description Index demonstrated significant convergent and discriminant validity. In addition, it showed less halo and central tendency, and had a higher information content than the other instruments. Implications of the findings are discussed.

Although the use of personality traits as criteria for performance appraisal is still dominant, behavior oriented measures of job performance have become more and more popular [Campbell, Dunnette, Lawler, and Weick, 1970, p. 64]. One of the first techniques utilised in this field was Flanagan's Critical Incident method [Flanagan, 1954]. He defined the critical requirements of a job as those behaviors which are crucial in making a difference between doing a job effectively and doing it ineffectively. This technique was expanded and modified by Smith and Kendall [1963] and Campbell, et al. [1970], and is now better known as Behaviorally Anchored Rating Scale (BARS), or Behavior Expectation Scale (BES) [Zedeck and Baker, 1972; Bernandin, LaShells, Smith, and Alvares, 1976].

According to Dunnette [1966], the behavioral approach to performance appraisal reduces much of the ambiguity of trait ratings by reducing the performance construct to the job itself (see also Harari and Zedeck, 1972). This desirable characteristic of behavioral scales is empirically supported by Fogli, Hulin, and Blood's [1971] and Smith and Kendall's [1963] generally high interrater agreement ($r \geq .97$) on the scale value assigned to behavioral descriptions and by the conceptual unidimensionality of behavioral incidents within scales [Campbell, et al., 1970].

Blood [1974] points to several positive "spin-off" effects from the use of behaviorally oriented scales: (1) It extends the domain of evaluated performance. As Borman [1974] has demonstrated, when performance evaluation scales are developed at different levels in an organization (superior, peers, subordinates) it is possible to get different behavioral dimensions from each group. Using as many organizational perspectives as possible (e.g. even clients) would

considerably broaden the job behavior domain. (2) It helps to develop training programs (see also Campbell, Dunnette, Arvey, and Hellervik, 1973). Since skills required on the job are described in specific behavioral terms--as compared to the name of a skill domain--shortcomings are much easier identified and specific training programs can be developed. In addition, since behavior description can be stated in negative ways, trainers are able to specify unwanted behaviors trainees should learn to avoid. (3) It enables assessment of agreement on organizational policy. This is possible because, for scale development, behavioral items are rated according to the level of performance they describe. Items with a large variance are usually discarded. However, these items may be the most important ones for organizational development, because they may pinpoint areas of organizational policy which are unclear, or may even point to areas of conflict. (4) It helps to assess the accuracy of communication of organizational policy. The mean item ratings can be collected from members of different organizational levels; e.g. foremen and workers, and compared as to their agreement. If both groups disagree on the desirability of specific behaviors, then obviously there is a communication problem which should be corrected.

Despite the intuitive appeal of behavior oriented scales [Bernardin, LaShells, Smith, and Alvares, 1976; Bernardin and Walter, 1977; Zedeck, Jacobs, and Kafry, 1976], and other suggested advantages [Zedeck, Imperato, Krausz, Oleno, 1974], recent reviews of the literature [Schwab, Heneman, and DeCotiis, 1975; Schwind, 1977] have not been encouraging. By comparing behavior oriented scales with Graphic Rating Scales (GRS) and/or Summated Rating Scales (SRS), researchers found either no or only modest superiority of BARS over GRS or SRS and, because of that,

questioned the justification for the investment of the vast amount of time and effort required for the development of BARS [Borman and Dunnette, 1975].

One of the major shortcomings of BARS seems to be the waste of valuable information [Schwab, et al., 1975; Schwind, 1977]. After the development or collection of critical incidents for a certain job, they are put through the validation, retranslation, and rating process, and they must fulfill the standard deviation criterion [Smith and Kendall, 1963]. Usually 20 to 50 critical incidents per job dimension survive. Yet only between 5 and 10, depending on the number of anchoring points of the scale, are utilised, all others are discarded. Undoubtedly those items which are not used contain valuable information about the job dimension to which they were attributed in the retranslation process. The decision to eliminate them is made on the basis of arbitrarily chosen criteria: a convenient mean value to fit the scale points, and the degree of agreement between raters as measured by the standard deviation.

A New Performance Evaluation Scale: The Behavior Description Index (BDI)

In order to improve the rating characteristics of behavior oriented scales and reduce their shortcomings, a new scale is proposed, based on descriptions of critical job behavior incidents. The development of these incidents is similar to the development of incidents for BARS (write-up, validation, retranslation) but the rating of the incidents from very ineffective to extremely effective is dropped. Instead of selecting 5, 7 or 9 incidents per job dimension (depending on the number of scale points chosen) the selection of a larger number of

behavioral incidents is suggested, the limit being determined only by the capacity--and willingness--of raters to read through and check off a certain amount of behavior descriptions. Tests so far have shown that raters are quite willing to check off 20 or 30 items per scale, with a total of 5 scales. It is expected that 30 or 40 items per scale may be optimal, since an increase in numbers of items very likely will trigger a fatigue effect in raters, resulting in such negative rating characteristics as halo and leniency.

The behavior incidents are listed randomly, each with a check-off column on the right side of the page (see Exhibit I). The rater responds to the question: "Does the ratee exhibit the described behavior consistently (all the time)?" or, in the case of an ineffective behavior sample: "Does the ratee exhibit the described behavior at all?" The responses (ratings) can either be evaluated individually to determine specific weaknesses, or can be converted into point scores. Of course, the responses can be given any weight the rater (or the evaluator of the ratings) desires. A simple solution would be to attribute one point to a positive response given to a positive item (a desirable behavior is exhibited consistently), similarly to a negative response given to a negative item (an undesirable behavior is not exhibited). A positive response to a negative item (a negative behavior is exhibited), and a negative response to a positive item (a positive behavior is not exhibited consistently) would result in zero points. A superior could determine in advance what scores would be acceptable or unacceptable; e.g. out of 30 possible points (assuming that 30 behavior samples were chosen per scale):-

0 - 15 may mean: urgent training required
(or, if measured after training; training ineffective)

16 - 20 may mean: training recommended

21 - 25 may mean: refresher course may be useful

26 - 30 may mean: no training required

Method

Development of Job Behavior Sample

Five large Canadian chartered banks were involved in the development and validation of the new scale. Since the use of the scale was planned primarily to be in the training evaluation area, it was agreed upon between the investigator and the 5 training managers of the banks to select as a target group assistant branch managers, since their job position required a large amount of continuous training. The training managers, themselves having been for at least two years in an assistant branch manager's position, determined that this job had 5 main aspects which could be described as:

1. Administration: quarterly analyses, monthly reports, routine work, branch security, internal correspondence, correspondence with other banks, co-operation with manager, supplies, housekeeping, repairs.
2. Customer Relations: complaints, correspondence, special service (counseling in complex matters), service maintenance, special efforts.
3. Marketing: advertising, display material, local campaigns, open house, marketing of bank services.
4. Personnel Administration: salary administration, performance appraisal, hiring and firing, disciplinary

actions, motivation, encouragement, moral support, personal problems, vacation and holiday scheduling.

5. Training: any job-related training and development activity, explanations, support in problem solving, advise, informing, up-dating, course scheduling, training follow-up.

Workshops were arranged with all 5 banks, attended by 5 branch managers, 5 assistant branch managers, and 5 bank clerks and tellers. The reason for this group mix was to take into account every aspect of the job of an assistant branch manager, as perceived by superiors, job incumbents, and their subordinates. Participants were informed about the purpose of the study, and asked to write down examples of effective and ineffective job behaviors of an assistant branch manager. Between 600 and 1,000 behavior samples were collected from each workshop. After editing and removal of redundant items, approximately 300 to 400 items per bank remained. They were listed in random order and submitted to groups of 5 judges (experienced job encumbants) from each bank. The judges were asked to determine whether each item was a valid description of an assistant branch manager's job behavior, and secondly, they had to categorise each item into one of the 5 job dimensions mentioned above (Administration, Customer Relations, Marketing, Personnel, and Training). Items were retained if 80% of the judges agreed on their validity and category. Approximately 80 to 150 behavior samples survived this process. These items were listed again, this time ordered into their respective categories, and sent to a second group consisting of 25 judges (in all banks). These judges were asked to rate each item on a 1 to 7 scale as to the degree of effectiveness the item described, ranging from very

very ineffective to extremely effective. The means of all ratings were calculated, and items were retained if the standard deviation of their ratings did not exceed 1.0. Approximately 50 to 80 behavior samples per dimension remained from each bank's initial sample pool. Those items were used to develop Behaviorally Anchored Rating Scales for each bank. Since the main purpose of the study was to develop an evaluation instrument which was valid for all banks, the validation process was not yet completed.

The remaining items from all banks were pooled according to their category, and redundant items removed. From the remaining behavior samples 50 items per job dimension were selected according to the lowest standard deviation criterion (i.e. highest degree of agreement among judges). These items were listed and sent to 200 judges in each bank, a total of 1,000 judges. They were asked to judge each item as to its validity as a typical job behavior of an assistant branch manager, and secondly, to rate it on a 1 to 7 scale as to the degree of effectiveness it described in their opinion. Out of 1,000 mailed rating lists, 511 usable were returned (response rate 51%). Items were retained if they fulfilled the following requirements:

1. 80% of <u>all</u> judges and
2. 60% of the judges <u>of the individual banks</u> agreed on the validity of the item;
3. the standard deviation of the item did not exceed 1.5.

A total of 120 items could be retained.

Development of Different Evaluation Instruments

In line with Schwab, et al.'s [1975] recommendation that for comparison purposes more than two instruments should be utilised, a Graphic Rating Scale (GRS), a Behaviorally Anchored Rating Scale (BARS), and a Behavior Description Index (BDI) were developed. The GRS was basically the adaptation of a 7-point rating scale used by the banks for their regular annual performance appraisal (see Exhibit 2). The BARS and the BDI instruments, on the other hand, were developed by using the job behavior samples mentioned before. The items for the BARS instrument were selected according to a) ratings closest to a scale point (1 to 7), and b) lowest possible standard deviation (see Exhibit 3). As it turned out, no items could be found for the midpoint range of the scales (4). This problem is discussed by Landy and Guion [1970], who suggest that polarisation is a necessary consequence of eliminating items which do not have meaning to individuals.

For the BDI instrument it was arbitrarily decided to utilise 20 job behavior samples per job dimension, or 100 in total. Examples of effective and ineffective behavior were selected according to the lowest standard deviation and listed in random order (see Exhibit 1).

As to the rating procedure, it was decided to adapt Smith, Kendall, and Hulin's [1969] approach to rate the responses to their measure of job satisfaction, the Job Description Index. Raters may respond "Yes" (Y), "No" (N), or "Undecided" (?). If a rater responds positively to an effective behavior sample, or negatively to an ineffective sample, the score will be 3 points. A positive response to a negative statement and vice versa results in 0 points. If the respondent is not sure or cannot decide, the response is a question mark, and the score will be 1 point.

In the latter case it is assumed that when a rater is uncertain whether a ratee exhibited a certain behavior or not it indicates a somewhat lower probability that the ratee will exhibit the desirable behavior (or not exhibit undesirable behavior). This is in line with the empirical findings by Smith, et al. [1969], who concluded that an "uncertain" response was more indicative of a negative response than a positive one, (p. 79).

Ratee and Rater Sample

This paper reports the results of the first test of the new instrument out of a series of several ongoing tests. A group of 25 assistant branch managers of one of the five Canadian banks were rated by their superiors, using the GRS, BARS, and the BDI as appraisal instruments. The sample was drawn from branches in three Maritime provinces, Nova Scotia, Prince Edward Island, and Newfoundland.

Analyses

Leniency will be measured by assessing the skewness of the different scales [Barrett, Taylor, Parker, and Martens, 1958; Haridas, Frost, and Barnowe, 1977], using a formula suggested by Blalock [1972]. The larger the skewness coefficient the larger is the leniency effect.

Halo effect will be assessed by averaging the correlations between ratings on all job dimensions [Barrett, et al., 1958].

Central Tendency will be measured by analysing the distribution of the scores for every instrument.

Information Content will be assessed by calculating the percentage of actually used job behavior samples from the total available item pool which survived the last validation process.

Construct Validity will be tested by using Campbell and Fiske's [1959] multitrait-multimethod matrix (15 scales x 3 methods).

Results

The means and standard deviations of the ratings are shown in Table 1. Since the BDI utilises a 60-point scale and the GRS and BARS a 7-point sacle, the standardised skewness coefficient for each performance dimension scores of the three instruments are shown in Table 2. An examination of the three skewness coefficients for each dimension reveals that the value for the BDI and the BARS instrument are similar, and closer to zero than the result for the GRS instrument.

The correlations between the different performance dimensions are reported in Table 3. The GRS scores show a significantly higher halo effect than both BARS and BDI scores ($p < .05$, for the GRS-BARS and GRS-BDI differences, according to the sign test; the difference between BARS and BDI is not significant). The GRS format yields 9 significant correlations out of 10, versus 3 for both the BARS and BDI format. By comparing the mean correlation for each instrument, the BDI comes out best with .24, as compared to .32 for the BARS and .49 for the GRS.

In order to measure central tendency effect of the three instruments the score distribution for the five dimensions of each instrument is shown in Table 4. It is interesting to see that the GRS scores contain no ratings in the two lowest scale ranges ("unacceptable" and "moderate"), and in the highest scale range ("excellent"). For the BARS and the BDI instruments only the lowest scale ranges is not utilised.

The degree of utilisation of available information is measured by calculating the percentage of actually used behavior samples from the total available item pool which survived the last validation process.

The results are shown in Table 5. The information content for the GRS could not be determined since it does not use job behavior samples. The data show that the BDI utilised almost three times as many job behavior samples as the BARS, or 83 percent as compared to 30 percent for the latter.

The final mode of analysis is the multitrait-multimethod approach suggested by Campbell and Fiske [1959]. The present study resulted in a 15 x 15 multitrait (5 performance dimensions) multimethod (GRS vs. BARS vs. BDI) matrix, shown in Table 6. Campbell and Fiske define convergent validity as the observation of significant correlations when two different methods are used to measure the same variables. As the results show, convergent validity is significant at the $p = .05$ and .01 level for all entries but the "Training" scale in the GRS/BARS correlations.

Discriminant validity is indicated in two ways. First, the entries in the validity diagonal can be compared with their corresponding row and column entries in the heterotrait-heteromethod triangles (indicated by broken lines). This yields 8 comparisons for each performance factor in which the diagonal value should be higher than the row and column values, if discriminant validity is present. Discriminant validity is highest for the BARS/BDI correlations, yielding 39 (out of 40 comparisons, or 97.5 percent) with higher validity entries, as compared to 36 for the GRS/BDI (90 percent), and 31 for the GRS/BARS correlation (77.5 percent). Since both convergent and discriminant validity are established for the BARS and BDI instruments, it can be concluded that both show evidence of construct validity. The GRS format compares less favorably.

A second index of discriminant validity involves comparing the validity diagonal entries (same traits but different methods) to the corresponding row and column entries in the heterotrait-monomethod triangles (indicated in solid lines). This implies that the correlations should be higher when the different methods are used to measure the same dimension than when different dimensions are measured by the same method. Discriminant validity according to the latter criterion is highest for the BARS and BDI instruments, yielding 17 (out of 20 comparisons) with higher validity entries, as compared to only 12 for the GRS instrument. A chi square test of independence indicates significantly fewer exceptions for the BARS and BDI instruments as compared to the GRS format at the $p < .01$ level ($\chi^2 = 8.64$; $df = 1$).

A further question concerns the pattern of relationships among traits when measured by the three rating methods. The ranks of the correlations in the three monomethod triangles were compared, resulting in $r = .31$ between GRS and BARS, $r = .35$ between GRS and BDI, and $r = .85$ between BARS and BDI. Only the latter correlation is strongly significant, the common variance being .72, suggesting a related structure between the performance dimensions of the two instruments.

Conclusions

From the results of the study several--cautious--conclusions can be drawn. First, behaviorally oriented scales need not be organization specific. This is an important finding, similar to the one by Goodale and Burke [1975] who demonstrated that BARS need not be job specific. The extension of the use of behavior oriented scales to other, but similar, jobs in the same organization and now even to similar jobs in

different organizations in the same industry should make the significantly higher investment in time and effort for the development of behavior oriented scales more worthwhile. It is now possible for organizations to pool their resources to develop such instruments, or delegate such a job to an institution to which all related organizations belong as members; e.g. the Institute of Canadian Bankers, the Trust Companies Association of Canada, the Canadian Manufacturers Association, and others.

As a second result of the study, it has been shown that it is possible to improve the psychometric characteristics of behavior oriented scales in several ways. One of the major shortcomings of the BARS format was the severely limited utilisation of the total pool of information available on the job behavior domain. The switch from a Likert-type scale--which limited the number of behavior samples used per scale to the number of anchor points--to an anchorless format which allows theoretically the use of the total job behavior domain, seems to be a breakthrough in the search for a performance appraisal instrument with a large and specific amount of feedback information and a superior rating characteristic than BARS.

And thirdly, because of the favorable rating characteristics, the increased utilisation of job information, and the very specific feedback to raters, evaluators, and ratees, it is possible that the BDI can be used as the basis or at least a significant part of a new Personnel Management System, utilising similar types of information in different areas. BDI-type of information can be used for:

1. Job Descriptions, to define scope of activities and major responsibilities connected with a job;
2. Job Analyses, to determine the behavioral aspects of a job;
3. Job Specification, to describe required skills to perform a specific job adequately;
4. Selection Aid, to check whether a "skilled" applicant actually possesses the skills for a specific job;
5. Determining Training Needs; no other instrument appears to be more suitable to assess an employee's strengths and weaknesses than the BDI, because no other device offers such specific feedback on areas where training is needed;
6. Assessment of Trainer's Effectiveness, to measure the specific strengths and weaknesses of instructors by focusing on their effective and ineffective teaching behavior;
7. Measuring Training and Development Program Effectiveness, to assess in detail the outcomes of all types of programs with the aim of changing a participant's behavior;
8. Performance Appraisals, to determine an employee's strengths and weaknesses (shortcomings) in the job behavior domain. The appraisal can be used for different purposes, such as:
   a) counseling, e.g. where to improve
   b) promotion, e.g. as evidence that required skills are demonstrated,
   c) wage and salary administration, e.g. rewards could be offered to those whose job behavior fulfills certain basic standards, and

9. Assessment of Agreement on Organizational Policy, to measure the accuracy of communication of organizational policies; e.g. desirability of specific behaviors, specially at different organizational levels.

It appears that the new behavior oriented evaluation scale, the Behavior Description Index, may be a needed addition to, or perhaps a replacement for, some of the instruments or processes used so far in the area of Personnel Management. Only future research can tell, how strong its position will be.

References

Barrett, R.S., Taylor, E.K., Parker, J.W., and Martens, W.L. "Rating scale content: 1. Scale information and supervisory ratings," Personnel Psychology, 1958, 11, 333-346

Bernardin, H.J., LaShells, M.B., Smith, P.C., and Alvares, K.M. "Behavioral expectation scales: Effects of developmental procedures and formats," Journal of Applied Psychology, 1976, 61, 75-79

Bernardin, H.J., and Walter, C.S. "Effects of rater training and diary-keeping on psychometric error in ratings," Journal of Applied Psychology, 1977, 62, 64-69

Blood, Milton R. "Spin-offs from behavioral expectation scale procedures," Journal of Applied Psychology, 1974, 59, 4, 513-515

Borman, W.C. "The ratings of individuals in organizations: An alternative approach," Organizational Behavior and Human Performance, 1974, 12, 105-124

Borman, W.C., and Dunnette, M.D. "Behavior-based vs. trait-oriented performance ratings: An empirical study," Journal of Applied Psychology, 1975, 60, 561-566

Campbell, J.P., Dunnette, M.D., Arvey, R.D., and Hellervik, L.W. "The development and evaluation of behaviorally based rating scales," Journal of Applied Psychology, 1973, 15-22

Campbell, J.P., Dunnette, M.D., Lawler, E.E., and Weick, K.E. Managerial Behavior, Performance, and Effectiveness. New York: McGraw-Hill Book Company, 1970

Campbell, D.T., and Fiske, D.W., "Convergent and discriminant validities by the multitrait - multimethod matrix," Psychological Bulletin, 1959, 56, 81-95

Dunnette, M.D. Personnel Selection and Placement. Belmont, California: Wadsworth, 1966

Flanagan, J.C. "The critical incident technique," Psychological Bulletin, 1954, 51, 327-358

Fogli, L., Hulin, C.L., and Blood, M.R. "Development of first-level behavioral criteria," Journal of Applied Psychology, 1971, 55, 3-8

Goodale, J.G., and Burke, R.J. "Behaviorally based rating scales need not to be job specific," Journal of Applied Psychology, 1975, 60, 389-391

Harari, O., and Zedeck, S. "Development of behaviorally anchored scales for the evaluation of faculty teaching," Journal of Applied Psychology, 1973, 58, 261-265

Haridas, T.P., Frost, P.J., and Barnowe, J.T. "Performance dimensions of teaching behavioral sciences in a Canadian university," Proceedings of the Administrative Sciences Association of Canada, 1977

Schwab, D.P., Heneman, H.G., and De Cotiis, T.A. "Behaviorally Anchored Rating Scales: A Review of the Literature," Personnel Psychology, 1975, 28, 549-562

Schwind, H.F. "New ways to evaluate teaching and training effectiveness," Paper presented at the Adult Education Research Association meeting in Minneapolis, April 1977

Smith, P.C., and Kendall, L.M. "Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales," Journal of Applied Psychology, 1963, 47, 149-155

Smith, P.C., Kendall, L.M., and Hulin, C.L. The Measurement of Satisfaction in Work and Retirement. Chicago: Rand, McNally and Co., 1969

Zedeck, S., and Baker, T. "Nursing performance by behavioral expectation scales: A multitrait - multirater analysis," Organizational Behavior and Human Performance, 1972, 7, 457-466

Zedeck, S., Imperato, N., Krausz, M., and Aleno, T. "Development of behaviorally anchored rating scales as a function of organizational level," Journal of Applied Psychology, 1974, 59, 249-252

Zedeck, S., Jacobs, R., and Kafry, D. "Behavioral expectation: Development of parallel forms and analysis of scale assumptions," Journal of Applied Psychology, 1976, 61, 112-115

Exhibit 1

Job Part: Administration

Description of Responsibilities: quarterly analyses, monthly reports, routine work, branch security, internal correspondence, correspondence with other banks, cooperation with manager, supplies, housekeeping, repairs

Rating Procedure: Please, THINK CAREFULLY before you answer the question: Does this job incumbent exhibit the in following described behavior all the time? (Does it describe his typical behavior?) Use "Y" for Yes, "N" for No, and "?" if you are not certain.

Delegates work to subordinates effectively by taking care that it is evenly distributed according to ability and knowledge of each employee.	
Has to borrow frequently required stationaries from other branches resulting in delays and disruption of work flow.	
Is unable to organize his own work load, is constantly "catching up" thus setting a poor example for others.	
Is very conscious of security, and effectively controls procedures thereby resulting in all personnel observing rules and regulations.	
Ensures that housekeeping of banking are is neat and desks are not messy, thus making the Branch look presentable and a pleasant place to do business.	
Drops work on a subordinate's desk to be completed without reviewing it with that particular individual, causing confusion and backlog.	
Has a high standard of letter writing, good typing, and ensures that letters are not signed by junior personnel.	
Keeps meticulous records and encourages staff to do the same, thereby reducing errors and rework.	
Keeps close control on cash holdings thus minimizing risks and maximizing profits.	
Offers suggestions to superiors in effort to improve effectivensss of new or existing policies.	
Is unable to develop a good diary system, causing confusion and delays with work.	
Does not care that correspondence is answered daily.	
Knows a new procedure well before explaining it to the staff.	
Keeps a check on amount of overtime and finds out reason.	
Analyzes the duties to be performed in order that matters of high priority are attended to first.	
Keeps down administrative costs through efficiency of operation, i.e., staff are profitably employed, over ordering of stationary is kept to a minimum, lights are not left on all night, staff do not work unnecessary overtime.	
Knows his manuals well and is able to quickly find procedures to follow.	
Keeps management informed of all aspects of branch operation.	
Takes proper safeguards to combat frauds, etc., such as holding funds until cheques clear, etc.	
Plans staff vacation so as to provide adequate customer service.	

Exhibit 2

Global Rating Scale

A team of training managers of 5 Canadian banks agreed that the job of a bank accountant (administration officer) can be divided into 5 areas: Administration, Customer Relations, Marketing, Personnel, and Training. You are asked to give an overall rating of the job performance of your accountant (administration officer) in the 5 areas mentioned, using a 7 point scale.

His performance is:

1	2	3	4	5	6	7
unacceptable	mediocre	slightly below average	average	slightly above average	superior	outstanding

Job Area	Please circle the appropriate number
Administration	1 2 3 4 5 6 7
Customer Relations	1 2 3 4 5 6 7
Marketing	1 2 3 4 5 6 7
Personnel	1 2 3 4 5 6 7
Training	1 2 3 4 5 6 7

Exhibit 3

Job Part: Administration

Description of Responsibilities: quarterly analyses, monthly reports, routine work, branch security, internal correspondence, correspondence with other banks, cooperation with manager, supplies, housekeeping, repairs

Rating Procedure: Read every example of the job incumbent's behavior and then put a check-mark (✓) by the example that best represents how you could expect the incumbent you are rating to typically perform in this aspect of his job.

This Accountant (Administration Officer):

7. ____ can be expected to be conscious of security, and effectively controls procedures, thereby resulting in all personnel observing rules and regulations.

6. ____ can be expected to keep down administrative costs through efficiency of operation, i.e., staff are profitably employed, over ordering of stationary is kept to a minimum, lights are not left on all night, staff do not work unnecessary overtime.

5. ____ can be expected to have the Current Account and Personal Chequing ledgers balanced daily.

3. ____ can be expected not to observe deadlines for reports therefore causing hasty jobs, often incorrect.

2. ____ can be expected to be unable to organize his own work load, constantly trying to "catch up", thus setting a poor example for others.

1. ____ can be expected to be unable to develop a good diary system, causing confusion and delays with work.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR EACH PERFORMANCE DIMENSION ON GRS, AOBARS, AND BDI RESPONSES

Performance Dimensions	GRS		AOBARS		BDI	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
A	5.16	.98	5.40	1.08	48.88	5.65
C	4.88	.78	5.50	.73	46.08	6.14
M	4.12	.78	4.94	1.19	33.80	10.55
P	4.84	.89	5.08	1.16	47.64	6.59
T	4.80	.81	5.50	.82	47.52	6.53

N=25

Note: A=Administration
C=Customer Relations
M=Marketing
P=Personnel
T=Training

GRS was scored on a 7 point scale
AOBARS was scored on a 7 point scale
BDI was scored on a 60 point scale

TABLE 2

NORMALIZED SKEWNESS COEFFICIENT FOR EACH PERFORMANCE DIMENSION ON GRS, AOBARS, AND BDI RESPONSES

	GRS	AOBARS	BDI
A	-2.75	-.55	- .26
C	- .76	1.50	-1.10
M	-1.90	-.78	-1.90
P	-1.01	-.91	.20
T	-2.22	.36	1.04
Mean	-1.72	-.82	-.90

N=25

Note: The mean skewness coefficients were calculated by using the absolute scores.

The skewness coefficients were calculated according to the formula

$$S_k = \frac{3\ (\ \bar{X} - Md\ )}{\sigma}$$ (Blalock, 1972)

$\bar{X}$ = Mean

Md = Median

σ = Standard deviation

TABLE 3

CORRELATIONS BETWEEN DIMENSION SCORES OF EACH OF THE THREE INSTRUMENTS

	GRS	AOBARS	BDI
A/C	51**	36	32
A/M	40*	18	-09
A/P	73**	62**	46*
A/T	61**	44*	51**
C/M	43*	07	00
C/P	38	38	32
C/T	41*	32	35
M/P	26	22	13
M/T	68**	06	-04
P/T	46*	56**	48*
Mean	49	32	24

N=25

A=Administration
C=Customer Relations
M=Marketing
P=Personnel
T=Training

*p< .05
**p< .01

TABLE 4

SCORE DISTRIBUTION FOR FIVE DIMENSIONS
OF EACH GRS, AOBARS, AND BDI INSTRUMENT

Scores	GRS	AOBARS	Scores	BDI
1	0	0	0-10	0
2	0	9	11-20	3
3	10	17	21-30	21
4	40	0	31-40	38
5	45	45	41-50	34
6	30	35	51-60	29
7	0	19		

N=25

TABLE 5

INFORMATION CONTENT FOR AOBARS AND BDI AS
MEASURED BY PERCENTAGE OF ACTUALLY UTILIZED
CRITICAL INCIDENTS FROM TOTAL AVAILABLE INCIDENT POOL

Total Available Pool	Utilized by AOBARS		Utilized by BDI	
#	#	%	#	%
120	30	25	100	83.3

TABLE 6

MULTITRAIT (PERFORMANCE DIMENSIONS), MULTIMETHOD (GRS vs. BARS vs. BDI) MATRIX FOR BANK ACCOUNTANTS

		GRS					BARS					BDI				
		A	C	M	P	T	A	C	M	P	T	A	C	M	P	T
GRS	A															
	C	51														
	M	40	43													
	P	73	38	26												
	T	61	41	68	46											
BARS	A	(50)**	20	08	36	14										
	C	34	(39)*	-18	37	03										
	M	61	32	(63)**	39	43										
	P	58	44	33	(64)**	25										
	T	63	54	19	25	(24)										
	A						36									
	C						18	07								
	M						62	38	22							
	P						44	32	06	56						
	T															
BDI	A	(54)**	13	-07	34	09	(81)**	36	22	41	28					
	C	46	(54)**	15	31	30	33	(58)**	16	37	52	32				
	M	18	17	(69)**	13	39	06	-41	(46)*	11	19	-09	00			
	P	59	13	00	(66)**	13	62	35	16	(69)**	40	46	32	13		
	T	71	29	-08	46	(45)*	36	35	23	29	(49)**	51	35	-04	48	

Note: N=25, ( )= Validity Diagonal, ◺ = Heterodimensional-Heteromethod Triangle, and ◺ = Heterodimensional-Monomethod Triangle; *p< .05, **p< .01